

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-023062

(43)Date of publication of application : 21.01.2000

(51)Int.Cl.

H04N 5/445

G06F 17/30

(21)Application number : 10-199517

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 30.06.1998

(72)Inventor : ABE SHOZO

MAEDA SHIGERU

ABE TETSUYA

OKITA HIDENORI

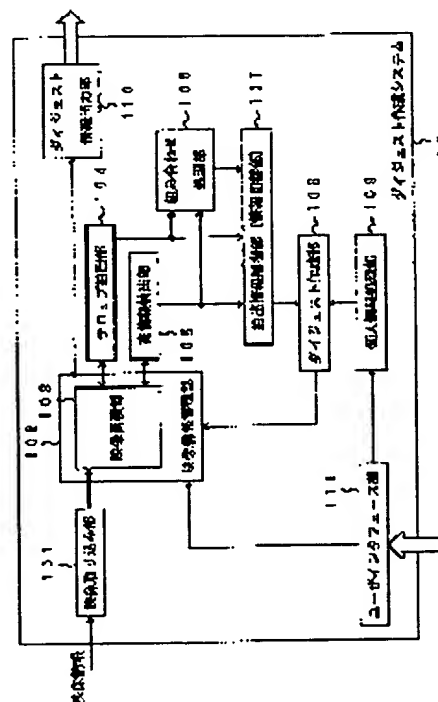
KONTA KAZUNOBU

## (54) DIGEST PRODUCTION SYSTEM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To efficiently produce a digest image from video information such as a video obtd. by recording a program based on audio information that synchronizes with the video information or a telop video in the video.

**SOLUTION:** A telop extracting part 104 extracts a telop video from a program video fetched by a video fetching part 101 and on the other hand, an audio information extracting part 105 extracts a characteristic image of a characteristic scene based on, e.g. a volume value of the audio information that is synchronized with the program video and is added. A combination processing part 106 extracts a video part that newly becomes characteristic due to the step wise combination of characteristics parts which correspond to the extracted telop video and audio information and a digest producing part 108 produces a digest video based on the extracted video.



BEST AVAILABLE COPY

## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

**\* NOTICES \***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.

2. \*\*\* shows the word which can not be translated.

3. In the drawings, any words are not translated.

**[Claim(s)]**

[Claim 1] The digest creation system characterized by to provide a sound information extract means extract the description image or the description scene of said image information, and a digest creation means create a digest image based on the description image or the description scene extracted by said sound information extract means, based on the sound information added synchronizing with image information.

[Claim 2] The digest creation system characterized by providing a telop extract means to extract a telop image including text from image information, and a digest creation means to create a digest image based on the telop image extracted by said telop extract means.

[Claim 3] A sound information extract means to extract the description image or the description scene of said image information based on the sound information added synchronizing with image information, A telop extract means to extract a telop image including text from said image information, The description image extracted by said sound information extract means or the description scene, and a combination processing means to extract the new description image combining the telop image extracted by said telop extract means, The digest creation system characterized by providing a digest creation means to create a digest image based on the description image extracted by said combination processing means.

[Claim 4] Said sound information extract means is the digest creation system according to claim 1 or 3 which is detecting the condition becoming beyond the setup time as which the time amount which becomes beyond the set point as which the sound-volume value of the sound information added synchronizing with said image information was determined beforehand continuously was determined beforehand, and is characterized by to extract the description scene of said image information in the description image or the time zone correspond of said image information at the time of corresponding.

[Claim 5] Said telop extract means is a digest creation system according to claim 2 or 3 characterized by extracting said telop image by detecting the change on the time-axis of the image data in the field concerned for the field of a proper in the class of the image concerned on an image screen, or the class of said telop image.

[Claim 6] Said image information set as the object of the telop image extract by said telop extract means is a program image. Said telop extract means A telop extract means for news flash to extract the telop image for news flash with checking change of the brightness information on a time-axis for the upper part on a program image screen, and a lower predetermined field, The change condition on the time-axis of color information is

checked for the whole program image screen. The digest creation system according to claim 5 characterized by having an emphasis telop extract means to extract an emphasis telop image by detecting the condition that the color extracted in the past on a time-axis and the left color became more than a setting ratio.

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention is used for the automatic image transcription system which is applied to the digest creation approach which creates a digest image from image information, especially records a program on videotape automatically, and relates to the suitable digest creation approach.

[0002]

[Description of the Prior Art] Some CS broadcasting is started besides a ground wave or BS broadcast in recent years, and viewing and listening of many programs is attained more. It will be expected from now on that the number of channels increases increasingly by digitization of broadcast. Thereby, the environment which chooses the program which suited its hobby idea from the channel which diversified the user more is ready.

Moreover, the technical environment of quality improvement including HD-TV is also being improved by many channelization and coincidence.

[0003]

[Problem(s) to be Solved by the Invention] However, it is becoming difficult to grasp whether there is any program he wants for covering to become difficult and to watch what kind of program is broadcast with the increment in the number of opposite side channels as a result. For this reason, there was a problem of overlooking without noticing what is being broadcast also in the program in which the user is interested.

[0004] When recording on videotape the program broadcast similarly, in order to have to choose a program to view and listen out of many programs, there was a problem that overlooking, a failure of an image transcription to take, etc. increased.

[0005] Thus, although there is [ that it is various and ] a problem in order to choose the program image transcription in alignment with an individual hobby idea towards the multi-channelization age, in order to watch the program which he wants to watch to the image which further once performed the program image transcription, the time amount which carried out the program image transcription had to be spent conventionally, or it had to carry out fast forwarding manually etc., and had to see from the beginning, and there was a problem took time amount. When the program was not a program which he wants to watch with interest, serious time amount was useless.

[0006] This invention was made in consideration of the above-mentioned situation, and the purpose is in offering the digest creation system which can create a digest image efficiently from image information, such as an image which carried out the program image transcription, based on the telop image in the sound information which synchronized with the image information concerned, or the image concerned.

[0007]

[Means for Solving the Problem] The digest creation system of this invention is characterized by to have a sound information extract means extract the description image or the description scene of image information, and a digest creation means create a digest image based on the description image or the description scene extracted by this sound

information extract means, based on the sound information added synchronizing with image information. In such a configuration, since the description image or the description scene of image information is extracted based on the sound information added synchronizing with image information, the digest image containing the description part relevant to sound information can be easily created by performing digest creation from image information using this description image or the description scene.

[0008] Here for the extract of the description image by the sound information extract means, or the description scene That what is necessary is just to detect the condition that the time amount (time zone) from which the sound-volume value concerned turns into beyond the set point continuously turns into beyond the setup time using the sound-volume value of the sound information added synchronizing with the above-mentioned image information For example, what is necessary is to extract the image at the initiation time of the condition concerned as a description image, or just to extract the sequential image in the condition (time zone) concerned as a description scene.

[0009] Moreover, it is also possible to classify the class of sound information other than the sound-volume value of sound information into voice, a sound effect, and sound musical sound, for example, for a user to specify, or to extract the image part corresponding to the description sound according to a user's hobby idea as a description image. Especially about voice, attribute discernment, such as an adult, a child, a male, and a woman, is performed. About a sound effect and music The image part corresponding to the description sound corresponding [ specify / a user / by preparing a corresponding sound pattern beforehand or ] to a user's hobby idea (for example, if a drive is the user of a hobby) The image scene corresponding to the transit sound of a car etc. is extracted on still finer level, and it can edit into the digest image according to a user's hobby idea by patching on a time-axis.

[0010] Moreover, this invention is characterized by having a telop extract means to extract a telop image including text from image information, and a digest creation means to create a digest image based on the telop image extracted by this telop extract means.

[0011] In such a configuration, since a telop image including text is extracted from image information, the digest image containing the description part relevant to a telop can be easily created by performing digest image creation using this telop image.

[0012] Here, if a telop image is extracted when the field on the image screen where a telop is displayed detects the change on the time-axis of the image data in the field concerned for the field in consideration of being the field of a proper in the class of the image, or the class of telop, an efficient telop image extract is realizable.

[0013] When the above-mentioned image information is a program image, it is necessary to identify two sorts, the telop for news flash used for a news program etc., and the emphasis telop used in a variety program etc. Both telops have the different description as the telop for news flash is displayed on the upper part of a screen, or a lower predetermined field and it is expressed as the color in which an emphasis telop is a big alphabetic character on the whole screen, and specification is conspicuous. Then, about the telop for news flash, it is checking change of the brightness information on a time-axis for the upper part on a program image screen, and a lower predetermined field, and when change is large, it is detectable as an initiation [ of the telop for news flash ], or termination point in time. On the other hand, about an emphasis telop, the change condition on the time-axis of color information is checked for the whole program image

screen, and it is detecting the condition the color extracted in the past on a time-axis and the left color having become more than a setting ratio, and can detect as an initiation [ of an emphasis telop ], or termination point in time.

[0014] Moreover, this invention establishes a combination processing means extract the new description image combining the description image extracted by the above-mentioned sound information extract means or the description scene, and the telop image extracted by the above-mentioned telop extract means, and is characterized by for the above-mentioned digest creation means to create a digest image based on the description image extracted by this combination processing means.

[0015] In such a configuration, it is creating a digest image combining two sorts of descriptions, and creation of the digest image of high quality is attained. The time amount range of the description scene which is the description image information for example, by telop display can be determined by using both the descriptions gradually especially, and the high digest image of quality can be created more by extracting the image scene from which a sound-volume value turns into beyond the set point by the time amount within the limits.

[0016] In addition, the more useful digest image for a user can be created by substituting the extracted image material for digest creation on a time-axis with the above-mentioned digest creation means.

[0017] Moreover, when treating the digitized image, the image information corresponding to the extracted description is acquired as a static image, and the large digest image of the application range can be created more by carrying out grouping according to each description conditions made into the purpose.

[0018]

[Embodiment of the Invention] Hereafter, with reference to a drawing, it explains per gestalt of operation of this invention.

[0019] Drawing 1 is the block block diagram of the information system equipped with the digest creation system concerning 1 operation gestalt of this invention.

[0020] The digest creation system 10 records TV program image on videotape on the VTR tape as a magnetic tape medium by TV program reservation system 20, and performs digest creation processing to the image (video image) recorded on the VTR tape, consequently outputs it to TV program reservation system 20 as a digest image.

[0021] TV program reservation system 20 cooperates with the digest creation system 10, and operates. TV program reservation system 20 performs image transcription processing of a program image in accordance with TV program which the user set up by considering the real image and TV program information (television race card) on TV (ground broadcast, satellite broadcasting service, etc.) as an input.

[0022] Efficient TV program reservation system becomes possible by using it combining the two above-mentioned systems 10 and 20.

[0023] For example, when a user looks at the image recorded on videotape in large quantities, it is inefficient to see over the same time amount as the time amount recorded on videotape, and it becomes useless [ time amount ]. Then, it is expected to the image recorded on videotape towards the many channels of the future digitization age that the need of creating a digest image becomes still stronger. In such times, the semantics which creates a digest image considers the significance of the need for a digest image of having been suitable for the so-called individual humanity news which included the individual

hobby idea in extent which the contents of the program image understand in addition to shortening in time.

[0024] Drawing 2 is the block diagram showing the configuration of the digest creation system 10.

[0025] In drawing 2, the image incorporation section 101 incorporates the image information recorded on videotape from the usual TV image, \*\*\*\*\*, cable television, etc. (with TV program reservation system 20 in drawing 1), and memorizes the whole program image in the image are recording section 103 which the image Research and Data Processing Department 102 has. The image Research and Data Processing Department 102 performs overall control management for creating a digest image to the image information accumulated in the image are recording section 103.

[0026] The telop extract section 104 is 1 functional element for digest image creation, and extracts the static image (frame image) taken out from the program image corresponding to the program image point (elapsed time which carried out counting from the beginning of a program image) with which the telop was displayed into the program image, and the continuous dynamic image of two or more sheets.

[0027] The sound information extract section 105 is 1 functional element for digest image creation like the telop extract section 104, and extracts the description part which summarizes a program image using the sound information which exists synchronizing with a program image.

[0028] The combination processing section 106 is used combining gradually the telop extract section 104 and the sound information extract section 105, and extracts a digest image with the description part of the program image taken out by telop extract, and the new description part with the description part of the program image taken out based on sound information.

[0029] The extract information control section (information change section) 107 changes suitably the telop extract section 104, the sound information extract section 105, and the combination processing section 106, and presents the digest image creation by the digest creation section 108 with them.

[0030] Using individual humanity news including the hobby idea of the individual who got by the individual humanity news processing section 109, the digest creation section 108 creates the final digest image based on the output of the extract information control section 107, and carries out transfer are recording in the image Research and Data Processing Department 102.

[0031] The digest information output section 110 outputs and displays the digest image accumulated in the image are recording section 103 at other systems (here TV program reservation system 20 in drawing 1).

[0032] The user interface section 111 makes an interface with a user, and performs the transfer to the individual humanity news processing section 109 of the individual humanity news which the user inputted, and the transfer to the image Research and Data Processing Department 102 of control information according to directions of a user.

[0033] Next, actuation of the digest creation system 10 constituted as mentioned above is explained.

[0034] First, the feature extraction (extract of the description point or the description scene) based on the sound information (specifically sound-volume value) by the sound information extract section 105 in the digest creation system 10 is explained with

reference to the explanatory view of drawing 3 of operation, and the flow chart of drawing 4 .

[0035] Here, with the sound information which synchronized with the program image concerned, the program image of the long sumo match recorded by the image transcription tape (VTR tape) TP on videotape by TV program reservation system 20 in drawing 1 for 2 hours shall be incorporated by the image incorporation section 101, and shall be accumulated in the image are recording section 103. The time amount range R1 in the image transcription tape TP shown in drawing 3 shows one measure.

[0036] What traced and graph-ized the sound-volume value in this R1 is shown in R2 region. In this measure, the frame image (or when Th1 is exceeded) of the beginning of a program image when becoming some one or more fixed sound-volume values (voice power level) Th defined beforehand is respectively extracted as an image of D1, D2, and D3.

[0037] The spectator was able to be excited as a measure situation, namely, if it thinks that it is shown that there were three scenes of the continuation time zone of the sound by the large cheer, it will become possible [ also extracting the program image corresponding to the time zone used as one or more constant value Th with sound volume continuously ]. When sound information is used, it can become a feature extraction with more effective extracting as a description scene which continued in time.

[0038] Then, the sound information extract section 105 in this operation gestalt When it is detected that checked the sound-volume value of the sound information which synchronized with the program image from the head of the image transcription tape TP (step S1), and the sound-volume value was set to one or more Th It investigates whether the periods (time zone) when whether the condition a sound-volume value being set to one or more Th continuing two or more fixed time amount Th beforehand defined from the point in time, and a sound-volume value that is, are continuously set to one or more Th are two or more Th (step S3). And the sound information extract section 105 is restricted when the condition that a sound-volume value is set to one or more Th is continuing two or more time amount Th, and it performs the next image extract.

[0039] The sound information extract section 105 judges first whether it extracts as a still picture, or it extracts as continuation drawing according to a user's assignment information that it was inputted into the image Research and Data Processing Department 102 through the user interface section 111 (step S4). If the still picture is specified, the sound information extract section 105 will extract the frame image of the head of a program image where the sound-volume value became one or more Th (step S5). (as the description point by the sound-volume value) On the other hand, if continuation drawing is specified, the sound information extract section 105 will extract the program image of the time zone when the sound-volume values are one or more Th (step S6). (as a description scene by the sound-volume value)

[0040] The sound information extract section 105 repeats the above actuation to the termination location of the program image recorded by the image transcription tape TP on videotape (step S7). That is, if it is the example of drawing 3 , the description point or the description scene by the sound-volume value will be extracted by performing processing about the time zone R1 of each measure in the image transcription tape TP for 2 hours of the whole of the image transcription tape TP concerned.



[0041] In addition, the monitor M1 on the actuation screen in drawing 3 is used for displaying an image transcription tape TP tape as it is, and a monitor M2 displaying the extracted description scene (digest image by the sound-volume description extracted by the sound information extract section 105 here).

[0042] Drawing 5 is drawing explaining how the telop extract section 104 extracts the image scene which shows the telop to the image which recorded similarly on videotape the image program of the long sumo match explained by drawing 3 on the image transcription tape TP for 2 hours.

[0043] In drawing 5, the time amount range R1 of the image transcription tape TP shows one measure. In the example of drawing 5, three scenes D11, D12, and D13, i.e., images, are extracted as a telop image within R1. The example of these images (telop image) D11, D12, and D13 is shown in drawing 6 (a), (b), and (c). The telop image of a waging-war partner's wrestler name with which the image D11 of drawing 6 (a) is displayed after a wrestler call, The image D12 of drawing 6 (b) the telop image of the past record-against-an-opponent table, and the image D13 of drawing 6 (c) The result of a measure, About the program image of the long sumo match which is the telop image of the winning technique at the time of receiving declaration of victory, since the pattern of a display telop is fixed, it can also extract R1 which is the one measure time amount range of the image transcription tape TP.

[0044] The extract of the telop image (screen) by the telop extract section 104 is performed as follows fundamentally. First, the number of pixels from which the difference information on the adjoining frame image in the direction of time amount becomes more than constant value about the whole image screen is counted, and it carries out by detecting the image with which the counted value became more than fixed numbers. In addition, difference information on the adjoining frame image in the direction of time amount is made binary by constant value about the whole image screen, and you may make it detect the image with which the frequency of occurrence of edge data becomes more than a certain constant value.

[0045] In examples, such as a previous long sumo match, the field on the screen where a telop is displayed is fixed like drawing 6. in such a case, it becomes possible to boil extract precision markedly and to make it high with setting up and checking a mask field beforehand. The field on the screen where a telop is displayed is prescribed by each program in many cases. It is also possible to set up the mask field of a proper for every following, for example, treating as one of the hobby idea information of a user, i.e., each user's favorite program.

[0046] Moreover, a much more efficient digest program image can be created by using combining the approach of the description image extract based on the extract of the description scene based on the sound-volume value by the sound information extract section 105, and the telop extract by the telop extract section 104 in the combination processing section 106. That is, if it is the example of a previous long sumo match, the time amount range of each measure can be determined for the description image information by telop display, and the high digest image of quality can be created more by extracting the image scene from which a sound-volume value turns into the one or more set points Th by the time amount within the limits.

[0047] When the telop extract section 104 extracts a telop image here and it extracts the emphasis telop image used abundantly in the detail of actuation in the case of creating a

digest image in the case where telop images for news flash, such as news, are extracted, a variety program, etc., it divides and explains.

[0048] Drawing 7 (a) shows the mask field set up in order to extract telop images for news flash, such as news. Here, the oblong rectangle field 31 of the screen upper part or the oblong rectangle field 32 of a bottom of screen is set up as a detection field (mask field) on the television screen 30, and the telop image for news flash is extracted with checking the temporal response of the image in the field 31 concerned or 32.

[0049] Hereafter, extract processing of the telop image for news flash by the telop extract section 104 is explained with reference to the flow chart of drawing 8.

[0050] First, the telop extract section 104 takes out the image frame before and behind on a time-axis from the head of the program image recorded by the image transcription tape TP on videotape one by one, and the difference information on a multiple value (subtraction image) which consists of difference of the pixels, for example, brightness data, of the same location of the mask field (31 or 32) of both the frame image is searched for (steps S11-S13).

[0051] The telop extract section 104 about the difference information (subtraction image) on the mask field (31 or 32) next, with a multiple-value image When the number of the pixels beyond a certain setting threshold is counted (step S15) and the counted value becomes more than the fixed (decided by total number of pixels of mask field) number (or fixed rate to the total number of pixels of a mask field) in the field concerned It is judged that telop images for news flash, such as news, were extracted (step S15).

[0052] Only by this decision technique, it becomes the extract only at the time of the time of a telop image being displayed and a telop display being completed. So, the following technique is applied with this operation gestalt.

[0053] Namely, if the time of a telop image being displayed is detected, the telop extract section 104 While holding the last frame image, i.e., the frame image before [ one ] displaying a telop, (steps S16 and S17) The flag (telop graphic display beginning flag) of the purport that the initiation time of a display of a telop image was detected is turned on (step S18), and the present frame image (here top telop image) taken out at step S12 is extracted as a telop image (step S19).

[0054] Henceforth, the telop extract section 104 takes out a consecutive frame image (step S12), and the difference information on the frame image before [ one ] displaying the telop held at step S17 is searched for (step S13). Thus, by searching for the difference information on the frame image before [ one ] displaying a telop, as long as the telop display continues, the number of the pixels beyond the setting threshold in difference information (subtraction image) serves as a big value more than the fixed number continuously (steps S14 and S15). Therefore, all the frame images in the period when it is in this condition, and (steps S16 and S19) and a telop are displayed by extracting the present frame image as a telop image when the telop graphic display beginning flag turns on can be extracted as a telop image.

[0055] If the number of the pixels beyond a setting threshold stops fulfilling the fixed number soon, the telop extract section 104 turns off a telop graphic display beginning flag (step S21), after holding the present frame image, will take out (step S22) and the following frame image (step S12), and will repeat the processing after the above mentioned step S13. when the number of the pixels beyond a setting threshold stopped

fulfilling the fixed number, in order [ thus, ] to return to the usual display -- difference -- the number of the pixels beyond the setting threshold in information decreases.

[0056] Drawing 7 (b) shows the mask field set up in order to extract the emphasis telop image currently used abundantly in the variety program etc. Here, as a detection field (mask field) on the television screen 30, the rectangle field 33 of the whole screen is set up mostly, and an emphasis telop image is extracted with checking the temporal response of the image of the field 33 concerned. It is in the point that a telop image is passed by the color which is a big alphabetic character on the whole screen as a description of an emphasis telop image, and is conspicuous.

[0057] therefore, this operation gestalt shows to the flow chart of drawing 9 -- as -- the extract of the previous telop image for news flash -- the same -- carrying out -- the difference on the time-axis of the image in a field 33 -- it processes (steps S21, S22-S24), and also the following processings are performed.

[0058] First, when the number of pixels beyond the setting threshold in the difference information (subtraction image) on a mask field (33) turns into more than the fixed number, the point (steps S26-S28) which holds the last frame image, i.e., the frame image before [ one ] displaying a telop, and turns on a telop graphic display beginning flag is the same as that of the case of an extract of the previous telop image for news flash. Differing from the extract of the telop image for news flash is the point (step S29) of extracting the period when the number of the pixels beyond the setting threshold in difference information (subtraction image) serves as a big value more than the fixed number continuously, and the present [ each time ] frame image at the time as an emphasis telop image candidate, and searching for the frequency information on the color in the frame image. In this case, by using a hue value as color information, it is stabilized and the extract of a color can be performed.

[0059] Now, if the number of the pixels beyond a setting threshold stops fulfilling the fixed number and returns to the usual display, the telop extract section 104 turns off a telop graphic display beginning flag, after holding the present frame image, it will measure the color frequency of (steps S30 and S31) and each emphasis telop image candidate who extracted till then, and will look for the color which serves as a fixed rate in the mask field (check field) 33 (step S32). And when the color which serves as a fixed rate in a field 33 exists, the telop extract section 104 judges the corresponding image to be an emphasis telop image, and memorizes it (step S34). When the color which serves as a fixed rate in a field 33 on the other hand does not exist, the telop extract section 104 is discarded noting that the corresponding image is not an emphasis telop image.

[0060] Here, as a digest image, the contents of the program image original with the image of only an emphasis telop image may be unclear. For this reason, after an emphasis telop image is displayed, it becomes possible to make the contents expressed by the emphasis telop and an image correspond by extracting the image of the time amount defined beforehand.

[0061] Next, how to substitute on a time-axis in the digest creation section 108 is explained with reference to drawing 10 using a user's individual humanity news to which the image material extracted to digest creation is given through the user interface section 111.

[0062] First, for example as a viewing-and-listening style of the cooking program image 40 of drawing 10 , when [ in a user's TV program ] the user concerned actually needs to

cook things, he wants to see the image of a cooking name or recipe information first in many cases, and a cooking process is referred to if needed.

[0063] However, with the digest image created by usual by technique which was already described, like drawing 10, the sequence of the description scene needs to serve as a list of the order of the telop image D21 of a "cooking program title", the telop image D22 of a "recipe", and the telop image D23 of "completion", and needs to look at all the time amount (here 10 minutes) of the cooking program image 40 as a user.

[0064] On the other hand, it arrange in order of the image D23 of the "completion" telop which substituted the time sequence of the extracted description scene like an arrow head 41 in consideration of a user's individual humanity news (liking), for example, displayed the cooking name by the digest creation section 108, and the telop image D22 which displayed "recipe" information, and the image scene of an actual cooking process be put in order after that. Even if the image scene of this cooking process extracts the thing of the section of the time amount L inserted by the telop image D22 of a "recipe", and the telop image 23 of "completion" as it is, and it samples in time, it is not cared about.

[0065] Although the above explained the digest creation system 10 used combining TV program reservation system 20, it does not restrict to this.

[0066] For example, as shown in drawing 11, it may be used, building the digest creation system 10 into the system (content provider side system) 50 which offers contents, such as TV program image. That is, it is also possible to use the digest creation system 10 for creating the digest image of the program image as which the content provider side system 50 was required from the user side system 60.

[0067] In the configuration of drawing 11, the user side system 60 requires the program image which the user itself wants to see of the content provider side system 50. Under the present circumstances, if it shall be constituted so that the system 50 concerned may send all the program images to the user side system 60 from the beginning when receiving offer of the demanded program image from the content provider side system 50, while transmitting an image with much capacity will take time amount, there is a problem whether it was the program image which a user wants to see truly.

[0068] On the other hand, in the video on-demand system of drawing 11, by the digest creation system 10 built into the content provider side system 50, the digest image of the program image which the user specified according to the demand of a user can be created, and it can provide for the user side system 60. Thereby, when the rough contents of the program image first made into the purpose with a digest image have been grasped, if a user is a program image to see truly, he should just demand a transfer of the original image of a program image again.

[0069] By the way, when a digest image is considered to be one epitome which shows the contents of the program image, it is not necessary to be the information which continued on the same time-axis towards digitization of the image spread is expected to be as an expression of the epitome of an image program which carried out clustering on the time-axis from now on. Using as an epitome which shows the contents of the movie is also possible by sticking on the movie introduction scenario map which extracts the movie contents for DVD as a static image in the image corresponding to the description information, and has prepared them beforehand there. Of course, if interested in the epitome, it is also possible to appreciate a original image.

[0070] Moreover, as applied with said operation gestalt, each point image on a time-axis by which it was characterized [ of the program image ] is extracted using telop information and sound information. By rearranging and editing into arbitration on a time-axis, without editing on the time-axis which the digest program which continued on the time-axis was created, and also original followed in each point image on the time-axis concerned. When appreciating a digest program, it makes it possible to see a result image immediately. As an effective example of edit by such rearrangement, application on sports program images, such as baseball and soccer, is. Here, the scene of seeing a game result first and seeing the contents slowly later can be considered.

[0071] The functional element group which constitutes the digest creation system 10 described above. Or the procedure performed in the procedure performed by the digest creation system 10 especially the telop extract section 104, and the sound information extract section 105. A computer is equipped with record media, such as CD-ROM by which the program for performing the procedure concerned was recorded on the program for operating a computer as a set of the functional element group, or the computer. It realizes by reading and performing the program concerned. This program may be loaded to a computer through communication media, such as a communication line.

[0072]

[Effect of the Invention] As having explain in full detail above , it be base on the approach processings , such as an information compression on a simple time-axis like \*\* of VTR equipment and image recognition , usually take time amount instantly to image information , especially the program image which recorded on videotape according to this invention , but since the scene and the scene used as the description of the program image concerned extract using the telop and the sound information expressing the attribute of a program image and it be consider as a digest image , the digest image concerned can edit flexibly .

[Brief Description of the Drawings]

[Drawing 1] The block diagram of the information system constituted combining TV program reservation system in the digest creation system concerning 1 operation gestalt of this invention.

[Drawing 2] The block diagram showing the configuration of the digest creation system 10 in drawing 1 .

[Drawing 3] Drawing for explaining the feature extraction based on the sound information (sound-volume value) by the sound information extract section 105.

[Drawing 4] The flow chart which shows the procedure of the feature extraction based on the sound information (sound-volume value) by the sound information extract section 105.

[Drawing 5] Drawing for explaining the telop image extract within the program image by the telop extract section 104.

[Drawing 6] Drawing showing an example of the telop image extracted from the program image of a long sumo match.

[Drawing 7] Drawing showing the mask field used for the telop image extract by the telop extract section 104 about each of the telop image extract for news flash, and an emphasis telop image extract.

[Drawing 8] Drawing showing the procedure of the telop image extract for news flash by the telop extract section 104.

[Drawing 9] Drawing showing the procedure of the emphasis telop image extract by the telop extract section 104.

[Drawing 10] Drawing for explaining the actuation which substitutes the image material extracted to digest creation on a time-axis by the digest creation section 108.

[Drawing 11] The block diagram showing the example of a configuration which applied the digest creation system concerning 1 operation gestalt of this invention to the video on-demand system.

[Description of Notations]

10 -- Digest creation system

20 -- TV program reservation system

50 -- Content provider side system

102 -- Image Research and Data Processing Department

104 -- Telop extract section

105 -- Sound information extract section

106 -- Combination processing section

107 -- Extract information control section

108 -- Digest creation section

109 -- Individual humanity news processing section

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2000-23062  
(P2000-23062A)

(43) 公表日 平成12年1月21日 (2000.1.21)

(51) IntCl.

識別番号

H 0 4 N 5/445

G 0 6 F 17/30

F I

H 0 4 N 5/445

G 0 6 F 15/40

15/401

テーマード(参考)

Z

3 7 0 D

3 2 0 Z

審査請求 未請求 請求項の数 6 F D (全 11 頁)

(21) 出願番号 特願平10-199517

(22) 出願日 平成10年6月30日 (1998. 6. 30)

(71) 出願人 060303576

株式会社東芝

神奈川県川崎市幸区船川町72番地

(72) 発明者 阿部 省三

神奈川県川崎市幸区柳町70番地 株式会社  
東芝柳町工場内

(72) 発明者 前田 茂

神奈川県川崎市幸区柳町70番地 株式会社  
東芝柳町工場内

(74) 代理人 100058479

弁理士 鈴江 武彦 (外6名)

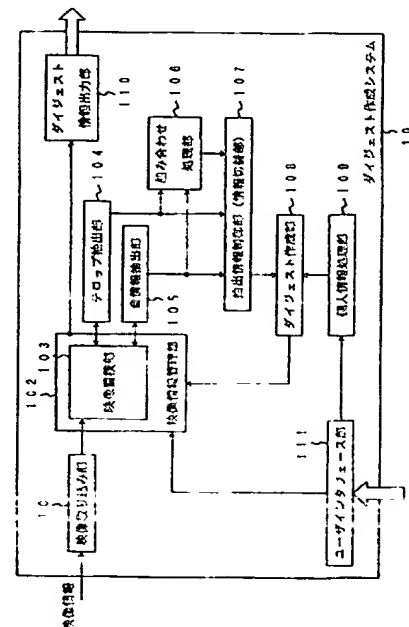
最終頁に続く

(54) 【発明の名称】 ダイジェスト作成システム

(57) 【要約】 (修正有)

【課題】番組録画した映像などの映像情報から、当該映像情報に同期した音情報、または当該映像中のテロップ映像をもとに、ダイジェスト映像を効率的に作成できるようにする。

【解決手段】映像取り込み部101により取り込まれた番組映像からテロップ抽出部104によりテロップ映像を抽出する一方、当該番組映像に同期して付加されている音情報の例えば音量値をもとに、音情報抽出部105により当該番組映像の特徴画像または特徴シーンを抽出する。組み合わせ処理部106は、抽出されたテロップ映像と音情報に対応した特徴部分の段階的組み合わせによる新たに特徴となる映像部分を抽出し、その抽出した映像をもとにダイジェスト作成部108にてダイジェスト映像を作成する。



## 【特許請求の範囲】

【請求項1】 映像情報に同期して付加されている音情報をもとに、前記映像情報の特徴画像または特徴シーンを抽出する音情報抽出手段と、

前記音情報抽出手段により抽出された特徴画像または特徴シーンをもとにダイジェスト映像を作成するダイジェスト作成手段とを具備することを特徴とするダイジェスト作成システム、

【請求項2】 映像情報から文字情報を抽出するテロップ映像を抽出するテロップ抽出手段と、

前記テロップ抽出手段により抽出されたテロップ映像をもとにダイジェスト映像を作成するダイジェスト作成手段とを具備することを特徴とするダイジェスト作成システム、

【請求項3】 映像情報に同期して付加されている音情報をもとに、前記映像情報の特徴画像または特徴シーンを抽出する音情報抽出手段と、

前記映像情報から文字情報を含むテロップ映像を抽出するテロップ抽出手段と、

前記音情報抽出手段により抽出された特徴画像または特徴シーン、及び前記テロップ抽出手段により抽出されたテロップ映像を組み合わせる新たな特徴映像を抽出する組み合わせ処理手段と、

前記組み合わせ処理手段により抽出された特徴映像をもとにダイジェスト映像を作成するダイジェスト作成手段とを具備することを特徴とするダイジェスト作成システム、

【請求項4】 前記音情報抽出手段は、前記映像情報に同期して付加されている音情報の音量値が連続して予め定められた設定値以上となる時間が予め定められた設定時間以上となる状態を検出することで、該当する時点における前記映像情報の特徴画像または該当する時間帯における前記映像情報の特徴シーンを抽出することを特徴とする請求項1または請求項3記載のダイジェスト作成システム、

【請求項5】 前記テロップ抽出手段は、映像画面上の当該映像の種類または前記テロップ映像の種類に固有の領域を対象に当該領域内の画像データの時間軸上の変化を検出することで前記テロップ映像を抽出することを特徴とする請求項2または請求項3記載のダイジェスト作成システム、

【請求項6】 前記テロップ抽出手段によるテロップ映像抽出の対象となる前記映像情報が番組映像であり、前記テロップ抽出手段は、番組映像画面上の上部及び下部の所定領域を対象に時間軸上の輝度情報の変化をチェックすることで連続テロップ映像を抽出する連続テロップ抽出手段と、番組映像画面全体を対象に色情報の時間軸上の変化状態をチェックし、時間軸上の過去に抽出した色と離れた色が設定比率以上となった状態を検出することで強調テロップ映像を抽出する強調テロップ抽

出手段とを備えていることを特徴とする請求項5記載のダイジェスト作成システム、

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、映像情報のダイジェスト映像を作成するダイジェスト作成方法に関し、特に放送番組を自動的に録画する自動録画システムに用いて好適なダイジェスト作成方法に関する。

## 【0002】

【従来の技術】近年、地上波やBS放送の他にケーブルCS放送が開始され、より多数の番組の視聴が可能になってきた。今後チャンネル数は放送のデジタル化によりますます増加することか予想される。これにより利用者は、より多様化したチャンネルから自分の趣味趣向にあった番組を選択する環境が整いつつある。また、多チャンネル化と同時にHDD-TVを始めとした高品質化の技術環境も整備されつつある。

## 【0003】

【発明が解決しようとする課題】しかし、反面チャンネル数の増加に伴いどのような番組が放送されているのかを網羅するのが困難となり、結果として自分の見たいと思う番組があるかどうかを把握することが難しくなってきた。このため、利用者が興味を持っている番組でも放送していることに気づかずに見逃してしまうなどの問題があった。

【0004】同様に放送された番組を録画する場合においても、多数の番組の中から視聴したい番組を選択しなければならないため、見逃しや録画の取り忘れなどが多くなるという問題があった。

【0005】このように、多チャンネル化時代に向けて、個人の趣味趣向に沿った番組録画を選択するためには、いろいろと問題があるが、更に、一旦番組録画を行った映像に対して、自分が見たい番組を見るためには、従来は番組録画した時間をかけたり、手動で早送りするなどして、また最初から見なければならず、時間がかかるという問題があった。もし、その番組が自分が興味を持って見たい番組でなかった場合には大変な時間の無駄となっていた。

【0006】本発明は上記事情を考慮してなされたものでその目的は、番組録画した映像などの映像情報から、当該映像情報に同期した音情報、または当該映像中のテロップ映像をもとに、ダイジェスト映像を効率的に作成することができるダイジェスト作成システムを提供することにある。

## 【0007】

【課題を解決するための手段】本発明のダイジェスト作成システムは、映像情報に同期して付加されている音情報をもとに、映像情報の特徴画像または特徴シーンを抽出する音情報抽出手段と、この音情報抽出手段により抽出された特徴画像または特徴シーンをもとにダイジェス



ト映像を作成するダイジェスト作成手段とを備えたことを特徴とする。このような構成においては、映像情報に同期して付加されている音情報をもとに、映像情報の特徴映像または特徴シーンを抽出されるので、この特徴映像または特徴シーンを利用して映像情報からのダイジェスト作成を行うことにより、音情報に同期した特徴部分を抽出したダイジェスト映像を簡単に作成することができ

る。  
【0008】ここで、音情報抽出手段による特徴映像または特徴シーンの抽出には、上記映像情報に同期して付加されている音情報の音価値を用い、当該音価値が連続して設定値以上となる時間（時間帯）が設定時間以上となる状態を検出すればよく、例えば当該状態の開始時点の画像を特徴映像として抽出するとか、当該状態（時間帯）における連続映像を特徴シーンとして抽出すればよい。

【0009】また、音情報の音価値の他に、音情報の種類を、音声、効果音、音楽音に分類して、例えばユーザの指定する、或いはユーザの趣味趣向に応じた特徴音に対応する画像部分を特徴映像として抽出することも可能である。特に、音声については、大人、子供、男性、女性といった属性識別を行い、効果音及び音楽については、対応する音パターンを予め用意しておくことで、ユーザの指定する、或いはユーザの趣味趣向に応じた特徴音に対応する画像部分（例えば、ドライブが趣味のユーザであれば、車両の走行音に対応した映像シーンなど）を一層細かなレベルで抽出し、時間軸上で切り張りすることで、ユーザの趣味趣向に応じたダイジェスト映像に編集できる。

【0010】また本発明は、映像情報から文字情報を含むテロップ映像を抽出するテロップ抽出手段と、このテロップ抽出手段により抽出されたテロップ映像をもとにダイジェスト映像を作成するダイジェスト作成手段とを備えたことを特徴とする。

【0011】このような構成においては、映像情報から文字情報を含むテロップ映像が抽出されるので、このテロップ映像を利用してダイジェスト映像作成を行うことにより、テロップに関連した特徴部分を含むダイジェスト映像を簡単に作成することができる。

【0012】ここで、テロップが表示される映像画面上の領域が、その映像の種類またはテロップの種類に固有の領域であることを考慮し、その領域を対象に当該領域内の画像データの時間軸上の変化を検出することによりテロップ映像を抽出するならば、効率的なテロップ映像抽出が実現できる。

【0013】上記映像情報か番組映像の場合には、ニュース番組等に用いられる速報用テロップと、バラエティ番組等に用いられる強調テロップの2種を識別する必要がある。速報用テロップは画面の上部または下部の所定領域に表示され、強調テロップは画面全体に大きな文字

で且つ特定の目立つ色で表示されるというように、両テロップは異なる特徴を有している。そこで、速報用テロップについては、番組映像画面上の上部及び下部の所定領域を対象に時間軸上の映像情報の変化をチェックすることで、変化の大きい時点を選報用テロップの開始時点または終了時点として検出できる。一方、強調テロップについては、番組映像画面全体を対象に色情報の時間軸上の変化状態をチェックし、時間軸上の過去に抽出した色と現在の色か設定値以上となった状態を検出することで、強調テロップの開始時点または終了時点として検出できる。

【0014】また本発明は、上記音情報抽出手段により抽出された特徴映像または特徴シーン、及び上記テロップ抽出手段により抽出されたテロップ映像を組み合わせて新たな特徴映像を抽出する組み合わせ処理手段を設け、この組み合わせ処理手段により抽出された特徴映像をもとに、上記ダイジェスト作成手段がダイジェスト映像を作成するようにしたことを特徴とする。

【0015】このような構成においては、2種の特徴を組み合わせてダイジェスト映像を作成することで、高品質のダイジェスト映像の作成が可能となる。特に、両特徴を段階的に使用することで、例えばテロップ表示による特徴映像情報である特徴シーンの時間範囲を決定し、その時間範囲内で、音価値が設定値以上となる映像シーンを抽出することで、より品質の高いダイジェスト映像を作成することができる。

【0016】この他に、抽出したダイジェスト作成用の映像素材を上記ダイジェスト作成手段にて時間軸上で差し替えることで、ユーザにとってより有用なダイジェスト映像を作成できる。

【0017】また、デジタル化した映像を扱う場合には、抽出した特徴に対応する映像情報を静止画像として取得し、目的とする各特徴条件によってグルーピングすることで、より応用範囲の広いダイジェスト映像を作成できる。

【0018】

【発明の実施の形態】以下、本発明の実施の形態につき図面を参照して説明する。

【0019】図1は本発明の一実施形態に係るダイジェスト作成システムを備えた情報システムのブロック構成図である。

【0020】ダイジェスト作成システム10は、TV番組予約システム20にてTV番組映像を磁気テープ媒体としてのVTRテープに録画し、そのVTRテープに記録された映像（ビデオ映像）に対してダイジェスト作成処理を行い、その結果、ダイジェスト映像としてTV番組予約システム20に出力するものである。

【0021】TV番組予約システム20は、ダイジェスト作成システム10と連動して動作する。TV番組予約システム20は、TV（地上放送、衛星放送など）のリ

アル映像とTV番組情報(テレビ番組表)を入力として、ユーザが設定したTV番組に沿って番組映像の録画処理を行うものである。

【0022】上記2つのシステム10、20を組み合わせて使用することによって、効率的なTV番組予約システムが可能となる。

【0023】例えば、大衆に録画した映像をユーザが見る場合に、録画した時間と同じ時間をかけて見るのは非効率であり時間的損失となる。そこで、従来のアナログ化時代の高チャンネルに向けて、録画した映像に対して、ダイジェスト映像を作成する必要性が益々強くなっていくと予想される。こういった時代において、ダイジェスト映像を作成する意味は、番組映像の内容が分かる程度に時間的に短くすることに加えて、個人の趣味趣向を含んだ、いわゆる個人情報に適したダイジェスト映像の必要性の重要度を考えている。

【0024】図2はダイジェスト作成システム10の構成を示すブロック図である。

【0025】図2において、映像取り込み部101は、(図1中のTV番組予約システム20にて)通常のTV映像、衛星放送、ケーブルテレビなどから録画した映像情報を取り込み、映像情報管理部102の持つ映像蓄積部103に番組映像全体を記憶する。映像情報管理部102は映像蓄積部103に蓄積した映像情報に対してダイジェスト映像を作成するための全体的な制御管理を行う。

【0026】テロップ抽出部104は、ダイジェスト映像作成のための1機能要素であり、番組映像中にテロップが表示された番組映像ポイント(番組映像の最初から計数した経過時間)に対応した番組映像から取り出された静止画像(フレーム画像)、及び複数枚の連続した動画像を抽出する。

【0027】音情報抽出部105は、テロップ抽出部104と同様にダイジェスト映像作成のための1機能要素であり、番組映像に同期して存在する音情報を使って番組映像を要約するような特徴部分を抽出する。

【0028】組み合わせ処理部106は、テロップ抽出部104及び音情報抽出部105を段階的に組み合わせ使用して、テロップ抽出により取り出される番組映像の特徴部分と、音情報に基づいて取り出される番組映像の特徴部分を持つ新たな特徴部分を持つダイジェスト映像を抽出する。

【0029】抽出情報制御部(情報切り替え部)107は、テロップ抽出部104、音情報抽出部105、及び組み合わせ処理部106を適宜切り替えてダイジェスト作成部108によるダイジェスト映像作成に供する。

【0030】ダイジェスト作成部108は、個人情報処理部109によって得た個人の趣味趣向を含んだ個人情報を使って、抽出情報制御部107の出力に基づく最終的なダイジェスト映像を作成し、映像情報管理部102

に転送蓄積する。

【0031】ダイジェスト抽出出力部110は、映像情報管理部103に蓄積されたダイジェスト映像を他システム(ここでは図1中のTV番組予約システム20)に出力・表示する。

【0032】ユーザインタフェース部111はユーザとのインタフェースをなすもので、ユーザの入力した個人情報、個人情報処理部109への転送、ユーザの指示に従う録画情報の映像情報管理部102への転送を行う。

【0033】次に、以上のように構成されたダイジェスト作成システム10の動作を説明する。

【0034】まず、ダイジェスト作成システム10内の音情報抽出部105による音情報(具体的には音量値)に基づく特徴抽出(特徴ポイントまたは特徴シーンの抽出)について、図3の動作説明図及び図4のフローチャートを参照して説明する。

【0035】ここでは、図1中のTV番組予約システム20にて録画テープ(VTRテープ)TPに2時間録画された大相撲の番組映像が、当該番組映像に同期した音情報と共に、映像取り込み部101により取り込まれて映像蓄積部103に蓄積されているものとする。図3に示す録画テープTP内の時間範囲R1は1つの取り組みを示している。

【0036】このR1内の音量値をトレースしてグラフ化したものをR2域に示す。この取り組みでは、予め定められたある一定の音量値(音声パワーレベル)Th1以上となったとき(あるいはTh1を越えたとき)の番組映像の最初のフレーム画像を各々D1、D2、D3の画像として抽出する。

【0037】取り組み状況としては、観客を沸かせた、即ち大歓声による音の連続時間帯の場面が3回あったことを示すと考えれば、音量がある一定値Th1以上となった時間帯に対応する番組映像を連続して抽出することも可能となる。音情報を使った場合には、時間的に連続した特徴シーンとして抽出する方が、より効果的な特徴抽出となり得る。

【0038】そこで本実施形態における音情報抽出部105は、録画テープTPの先頭から番組映像に同期した音情報の音量値をチェックし(ステップS1)、音量値がTh1以上となったことを検出した場合には、音量値がTh1以上となる状態が、その時点から予め定められた一定時間Th2以上継続するかどうか、つまり音量値が連続してTh1以上となる期間(時間帯)がTh2以上であるかを調べる(ステップS3)。そして音情報抽出部105は、音量値がTh1以上となる状態が時間Th2以上継続している場合に限り、次の映像抽出を行う。

【0039】まず音情報抽出部105は、ユーザインタフェース部111を介して映像情報管理部102に入力されたユーザの指定情報に従い、静止画として抽出するか連続画として抽出するかを判断する(ステップS

4)。もし、静止画が指定されているならば、音情報抽出部105は音量値がTh以上となった番組映像の先頭のフレーム画像を（音情報による特徴ポイントとして）抽出する（ステップS5）。これに対し、連続画が指定されているならば、音情報抽出部105は音量値がTh以上となっている時間帯の番組映像を（音情報による特徴シーンとして）抽出する（ステップS6）。

【0040】音情報抽出部105は以上の動作を録画タイプに記録された番組映像の終了位置まで繰り返す（ステップS7）。つまり、図3の例であれば、録画タイプTP内の各取り組みの時間帯R1についての処理を、当該録画タイプTPの2時間全体に対して行うことによって、音情報による特徴ポイントまたは特徴シーンを抽出する。

【0041】なお、図3中の操作画面上のモニタM1は録画タイプTPテープをそのまま表示し、モニタM2は抽出された特徴シーン（ここでは、音情報抽出部105により抽出された音量特徴によるダイジェスト映像）を表示するのに用いられる。

【0042】図5は図3で説明した大相撲の映像番組を同様に録画タイプTPに2時間録画した映像に対して、テロップを表示している映像場面をテロップ抽出部104により抽出する方法について説明する図である。

【0043】図5において、録画タイプTPの時間範囲R1は1つの取り組みを示している。図5の例では、R1内でテロップ映像として3箇所の場面、即ち画像D11、D12、D13を抽出している。この画像（テロップ画像）D11、D12、D13の具体例を図6

(a)、(b)、(c)に示す。図6(a)の画像D11は力士呼び出しの後に表示される対戦相手の力士名のテロップ画像、図6(b)の画像D12は過去の対戦成績表のテロップ画像、図6(c)の画像D13は取り組みの結果、勝ち名乗りを受ける際の決まり手のテロップ画像である大相撲の番組映像については表示テロップのパターンは固定化されているため、録画タイプTPの1つの取り組み時間範囲であるR1を抽出することも可能である。

【0044】テロップ抽出部104によるテロップ画像（画面）の抽出は、基本的には次のように行われる。まず、映像画面全体について時間方向での隣接するフレーム画像の差情報が一定値以上となる画素数をカウントし、そのカウント値が一定数以上となった画像を検出することで行う。なお、映像画面全体について時間方向での隣接するフレーム画像の差情報を一定値で2値化し、エッジデータの出現頻度が、ある一定値以上となる画像を検出するようにしても構わない。

【0045】先の大相撲などの例では、テロップが表示される画面上の領域は、図6のように固定化されている。このような場合には、予めマスク領域を設定してチェックすることで抽出精度を格段に高くすることが可能

となる。テロップが表示される画面上の領域は各番組によって規定される場合が多い。したがって、例えば、ユーザの趣味趣向情報の一つとして扱うこと、つまり各ユーザの好みの番組毎に、固有のマスク領域を設定することも可能である。

【0046】また、音情報抽出部105による音情報に基づく特徴シーンの抽出と、テロップ抽出部104によるテロップ抽出に基づく特徴映像抽出の方法を、組み合わせ処理部106にて組み合わせることで、より効果的なダイジェスト番組映像を作成することかてきる。つまり、先の大相撲の例であれば、テロップ表示による特徴映像情報と各取り組みの映像画面を決定し、その時間範囲内で、音量値が設定値Th以上となる映像シーンを抽出することで、より品質の高いダイジェスト映像を作成することかてきる。

【0047】ここでテロップ抽出部104によりテロップ映像を抽出して、ダイジェスト映像を作成する場合の動作の詳細を、ニュースなどの速報用テロップ映像を抽出する場合と、バラエティ番組などで多用される強調テロップ映像を抽出する場合に分けて説明する。

【0048】図7(a)は、ニュースなどの速報用テロップ映像を抽出するために設定されるマスク領域を示す。ここでは、テレビ画面30上での検出領域（マスク領域）として、画面上部の横長の矩形領域31、または画面下部の横長の矩形領域32を設定し、当該領域31または32内の画像の時間的変化をチェックすることで、速報用テロップ映像を抽出する。

【0049】以下、テロップ抽出部104による速報用テロップ映像の抽出処理について図8のフローチャートを参照して説明する。

【0050】まずテロップ抽出部104は、録画タイプTPに録画された番組映像の先頭から時間軸上の前後の映像フレームを順次取り出し、両フレーム画像のマスク領域（31または32）の同一位置の画素同士例えば輝度データの差分からなる多値の差情報（差分画像）を求める（ステップS11～S13）。

【0051】次にテロップ抽出部104は、そのマスク領域（31または32）の差情報（差分画像）について、多値画像のままで、ある設定閾値以上の画素の数をカウントし（ステップS15）、そのカウント値が当該領域内で（マスク領域の総画素数で決まる）一定個数（あるいはマスク領域の総画素数に対する一定の割合）以上となった場合に、ニュースなどの速報用テロップ映像が抽出されたと判断する（ステップS15）。

【0052】この判断手法だけでは、テロップ映像が表示される時点とテロップ表示が終了した時点のみの抽出となる。そこで本実施形態では、以下の手法を適用する。

【0053】即ちテロップ抽出部104は、テロップ映像が表示される時点を検出すると、直前のフレーム画

像、即ちテロップを表示する1つ前のフレーム画像を保持すると共に(ステップS16、S17) テロップ映像の表示の開始時刻を抽出した時のフラグ(テロップ映像表示開始フラグ)をONし(ステップS18)、ステップS12で取り出した現フレーム画像(ここでは、先頭のテロップ映像)をテロップ映像として抽出する(ステップS19)。

【0054】以後テロップ抽出部104は、現在のフレーム画像を取り出して(ステップS12)、ステップS17で保持しておいたテロップを表示する1つ前のフレーム画像との差分情報を求める(ステップS13)。このように、テロップを表示する1つ前のフレーム画像との差分情報を求めることで、テロップ表示が流れている限り、差分情報(差分画像)における設定閾値以上の画素の数は連続して一定個数以上の大きな値となる(ステップS14、S15)。したがって、この状態で、且つテロップ映像表示開始フラグがONしている場合には、現フレーム画像をテロップ映像として抽出することで(ステップS16、S19)、テロップが表示されている期間における全フレーム画像をテロップ映像として抽出できる。

【0055】やがて設定閾値以上の画素の数が一定個数に満たなくなると、テロップ抽出部104はテロップ映像表示開始フラグをOFFし(ステップS21)、現フレーム画像を保持した上で(ステップS22)、次のフレーム画像を取り出して(ステップS12)、前記したステップS13以降の処理を繰り返す。このように設定閾値以上の画素の数が一定個数に満たなくなった時点では、通常の表示に戻るため、差分情報における設定閾値以上の画素の数は少なくなる。

【0056】図7(b)は、バラエティ番組などで多用されている強調テロップ映像を抽出するために設定されるマスク領域を示す。ここでは、テレビ画面30上での検出領域(マスク領域)として、ほぼ画面全体の矩形領域33を設定し、当該領域33の画像の時間的変化をチェックすることで、強調テロップ映像を抽出する。強調テロップ映像の特徴としては、画面全体に大きな文字で、且つ目立つ色でテロップ映像が流される点にある。

【0057】したがって本実施形態では、図9のフローチャートに示すように、先の速報用テロップ映像の抽出と同様にして、領域33内の画像の時間軸上での差分処理(ステップS21、S22～S24)を行う他に、以下の処理を行う。

【0058】まず、マスク領域(33)の差分情報(差分画像)における設定閾値以上の画素数が一定個数以上となった時点で、直前のフレーム画像、即ちテロップを表示する1つ前のフレーム画像を保持し、テロップ映像表示開始フラグをONする点(ステップS26～S28)は、先の速報用テロップ映像の抽出の場合と同様である。速報用テロップ映像の抽出と異なるのは、差分情報

(差分画像)における設定閾値以上の画素の数が連続して一定個数以上の大きな値となっている期間。その都度、その時点における現フレーム画像を強調テロップ映像候補として抽出して、そのフレーム画像における色の強度情報を求めておく点(ステップS29)である。この場合、色情報として色相値を使用することで、色の抽出が容易に行える。

【0059】さてテロップ抽出部104は、設定閾値以上の色の数が一定個数に満たなくなり、一部の表示に戻ると、テロップ映像表示開始フラグをOFFし、現フレーム画像を保持した上で(ステップS30、S31)、それまで抽出しておいた各強調テロップ映像候補の色相値を比較して、マスク領域(チェック領域)33内で一定の割合となる色を探り(ステップS32)、そしてテロップ抽出部104は、領域33内で一定の割合となる色が存在する場合に、該当する映像を強調テロップ映像と判断して記憶する(ステップS34)。一方、領域33内で一定の割合となる色が存在しない場合には、テロップ抽出部104は該当する映像は強調テロップ映像ではないとして、廃棄する。

【0060】ここで、ダイジェスト映像としては、強調テロップ映像のみの映像では、本来の番組映像の内容が分かりにくい可能性がある。このため、強調テロップ映像が表示された後、予め定められた時間の映像を抽出することにより、強調テロップで表現した内容と映像を対応させることが可能となる。

【0061】次に、ダイジェスト作成用に抽出した映像素材をユーザインタフェース部111を通して与えられるユーザの個人情報を使って、ダイジェスト作成部108にて時間軸上で差し替える方法について、図10を参照して説明する。

【0062】まずユーザのTV番組における例えば図10の料理番組映像40の視聴スタイルとして、当該ユーザが実際に料理を作る必要がある場合には、料理名やレシピ情報の映像を最初に見たいことが多く、必要に応じて料理過程を参照する。

【0063】ところが、既に述べたような手法で通常に作成されるダイジェスト映像では、特徴シーンの順番は、図10のように、「料理番組タイトル」のテロップ画像D21、「レシピ」のテロップ画像D22、そして「出来上がり」のテロップ画像D23の順の並びとなり、ユーザとしては料理番組映像40の時間(ここでは10分)全部を見る必要がある。

【0064】これに対して、ダイジェスト作成部108により、ユーザの個人情報(好み)を考慮して、久印41のように、抽出した特徴シーンの時間的順番を差し替えて、例えば、料理名を表示した「出来上がり」テロップの画像D23、「レシピ」情報を表示したテロップ画像D22の順に並べ、その後実際の料理過程の映像シーンを並べる。この料理過程の映像シーンは、「レシ

ビ」のテロップ画像D22と「出来」かり」のテロップ画像23とて括まれた時間1.の区間のものをそのまま抽出しても、時間的にサンプリングしたものであっても構わない。

【0065】以上は、TV番組予約システム20と組み合わせて使用されるダイジェスト作成システム10について説明した。これに限るものではない。

【0066】例えば、図11に示すように、ダイジェスト作成システム10は、TV番組予約システム20とコンテンツを提供するシステム（コンテンツプロバイダ側システム50）50に組み込まれて使用されるものであっても構わない。つまり、コンテンツプロバイダ側システム50からユーザ側システム60から要求された番組映像のダイジェスト映像を作成するのにダイジェスト作成システム10を用いることも可能である。

【0067】図11の構成において、ユーザ側システム60は、ユーザ自身が見たい番組映像をコンテンツプロバイダ側システム50に要求する。この際、要求した番組映像の提供をコンテンツプロバイダ側システム50から受ける場合に、当該システム50が最初から番組映像の全てをユーザ側システム60に送るように構成されているものとする、容量の多い映像を転送するのに時間がかかると共に、ユーザが本当に見たい番組映像であったのかといった問題がある。

【0068】これに対して、図11のビデオ・オン・デマンドシステムにおいては、コンテンツプロバイダ側システム50に組み込まれたダイジェスト作成システム10により、ユーザの要求に応じてユーザの指定した番組映像のダイジェスト映像を作成して、ユーザ側システム60に提供することができる。これによりユーザは、まず最初にダイジェスト映像で目的とする番組映像の大まかな内容を把握した上で、もし本当に見たい番組映像であるならば、再度、番組映像の原映像の転送を要求すればよい。

【0069】ところで、ダイジェスト映像を番組映像の内容を示す1つの要約と考えた場合、今後、普及が予想される映像のデジタル化に向けて、時間軸上でクラスタ化した映像番組の要約の表現として、同様な時間軸上で連続した情報である必要はない。そこで、例えばDVD用の映画コンテンツをその特徴情報に対応した映像中の静止画像として抽出し、予め用意している映画紹介ナリオ・マップに貼り付けることによって映画の内容を示す要約として利用することも可能である。勿論、その要約に興味があれば、原映像を鑑賞することも可能である。

【0070】また、前記実施形態で説明したように、テロップ情報及び音情報を用いて、番組映像の特徴とした、時間軸上の各ポイント映像を抽出して、時間軸上で連続したダイジェスト番組を作成する他に、当該時間軸上の各ポイント映像を本来の連続した時間軸上に編集

ずに、時間軸上で任意に並べ替えて編集することによって、ダイジェスト番組を鑑賞する場合においても直ちに結果映像を見ることを可能とする。このような並べ替えによる編集の目的的な例として、野球やサッカーなどのスポーツ番組映像への応用がある。ここでは、試合結果を最初に見て、後でゆっくりと内容を見るといった場面を考えることができる。

【0071】以上に述べたダイジェスト作成システム10を構成する処理手順（あるいはダイジェスト作成システム10にて実行される処理手順、特にテロップ抽出部104及び音情報抽出部105にて実行される処理手順）は、コンピュータをその処理装置の集合として提供させるためのプログラム、あるいはコンピュータに当該処理手順を実行させるためのプログラムが記録されたCD-ROM等の記録媒体をコンピュータに装載して、当該プログラムを読み取り実行させることにより実現される。このプログラムが、通信回線等の通信媒体を通してコンピュータにロードされるものであってもよい。

【0072】

【発明の効果】以上詳述したように本発明によれば、映像情報、特に録画した番組映像に対して、通常VTR装置の早送りのような単純な時間軸上の情報圧縮、映像認識など処理に時間を要する方法によらず、番組映像の属性を表現しているテロップや音情報を使って、当該番組映像の特徴となる場面やシーンを抽出してダイジェスト映像としているため、当該ダイジェスト映像をフレキシブルに編集できる。

【図面の簡単な説明】

【図1】本発明の一実施形態に係るダイジェスト作成システムをTV番組予約システムと組み合わせて構成された情報システムのブロック図。

【図2】図1中のダイジェスト作成システム10の構成を示すブロック図。

【図3】音情報抽出部105による音情報（音量値）に基づく特徴抽出を説明するための図。

【図4】音情報抽出部105による音情報（音量値）に基づく特徴抽出の処理手順を示すフローチャート。

【図5】テロップ抽出部104による番組映像内のテロップ映像抽出を説明するための図。

【図6】大相撲の番組映像から抽出されたテロップ映像の一例を示す図。

【図7】テロップ抽出部104によるテロップ映像抽出に用いられるマスク領域を、速報用テロップ映像抽出と強調テロップ映像抽出の各々について示す図。

【図8】テロップ抽出部104による速報用テロップ映像抽出の処理手順を示す図。

【図9】テロップ抽出部104による強調テロップ映像抽出の処理手順を示す図。

【図10】ダイジェスト作成用に抽出した映像素材をダイジェスト作成部108により時間軸上で差し替える動

作を説明するための図。

【図11】本発明の一実施形態に係るタイジェスト作成システムをビデオ・オン・デマンドシステムに適用した構成を示すブロック図。

【符号の説明】

10…タイジェスト作成システム

20…TV番組予約システム

50…コンテンツプロバイダシステム

\*102…映像情報管理部

104…テロップ抽出部

105…音情報抽出部

106…組み合わせ処理部

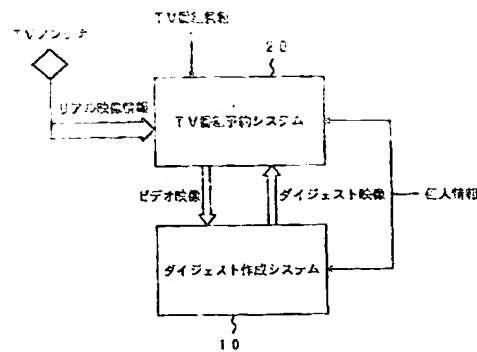
107…抽出情報制御部

108…タイジェスト作成部

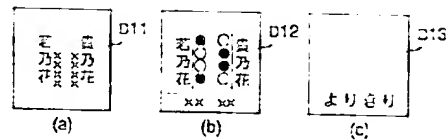
109…個人情報処理部

\*

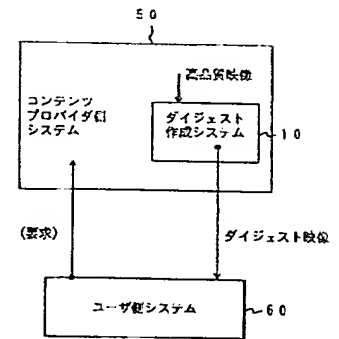
【図1】



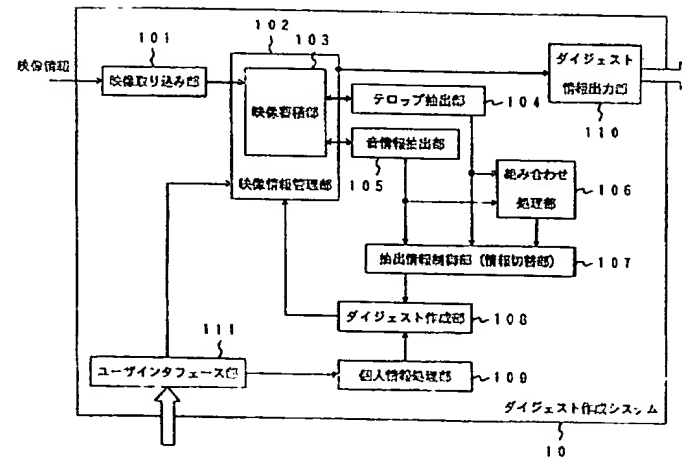
【図6】



【図11】

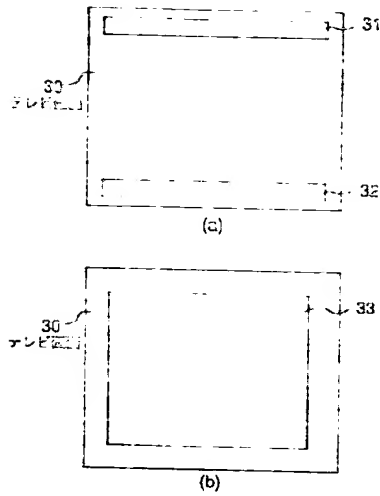


【図2】

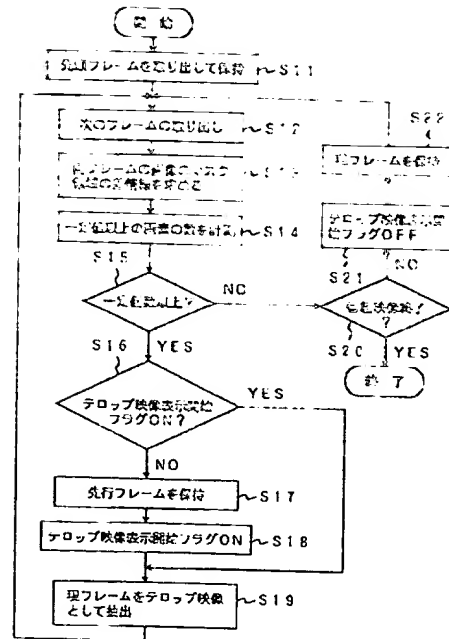




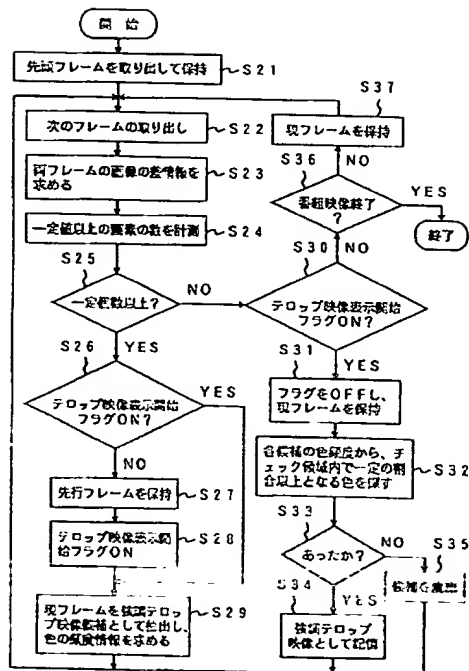
【図7】



【図8】

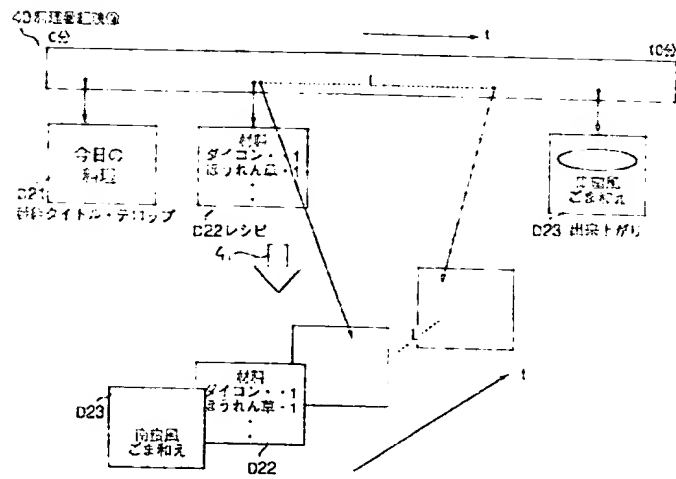


【図9】





【図10】



フロントページの続き

(72)発明者 阿部 哲也  
神奈川県川崎市幸区柳町70番地 株式会社  
東芝柳町工場内

(72)発明者 大高多 秀紀  
神奈川県川崎市幸区柳町70番地 株式会社  
東芝柳町工場内  
(72)発明者 紺田 和宣  
神奈川県川崎市幸区柳町70番地 株式会社  
東芝柳町工場内